

DOI:10.13995/j.cnki.11-1802/ts.022689

引用格式:蔡德玲,唐春华,梁玉英,等.融合近红外光谱和颜色参数的草莓可溶性固形物含量定量分析模型构建[J].食品与发酵工业,2020,46(7):218-224. CAI Deling, TANG Chunhua, LIANG Yuying, et al. Establishment of quantitative analysis model for detecting the soluble solids content in strawberry by merging near infrared spectroscopy and color parameters[J]. Food and Fermentation Industries, 2020, 46(7): 218-224.

融合近红外光谱和颜色参数的草莓可溶性固形物含量定量分析模型构建

蔡德玲^{1*}, 唐春华², 梁玉英¹, 曾川¹, 彭碧宁¹

1(中华人民共和国拱北海关技术中心, 广东 珠海, 519000) 2(珠海城市职业技术学院, 广东 珠海, 519090)

摘要 草莓可溶性固形物(soluble solids content, SSC)含量是评价草莓内部品质的关键指标。为了实现对该指标的快速、无损评估,基于近红外光谱技术,构建了线性偏最小二乘(partial least squares, PLS)和非线性最小二乘支持向量机(least squares support vector machine, LS-SVM)模型,联合蒙特卡罗无信息变量消除和连续投影算法(Monte-Carlo uninformative variable elimination, successive projections algorithm, MC-UVE-SPA)从原始光谱4 254个变量中提取了27个有效变量,并构建了基于有效变量的定量分析模型。同时,考虑到草莓表面颜色的影响,基于草莓RGB图像各分量获取了颜色特征参数,进一步融合光谱和颜色特征构建了多参数融合PLS和LS-SVM模型。基于相同的校正集和预测集,比较了所有模型对草莓内部SSC的预测性能。结果表明,MC-UVE-SPA是一种有效的草莓光谱变量选择算法,且多参数融合非线性LS-SVM模型是草莓内部SSC定量预测的最优模型。针对预测集样本,该模型相关系数 R_p 和预测均方根误差(root mean square error of prediction, RMSEP)分别为0.988 5和0.153 2。该研究为基于近红外光谱技术的草莓可溶性固形物含量检测便携式仪器和在线检测设备研发奠定了基础。

关键词 草莓;近红外光谱;可溶性固形物含量;参数融合;建模分析

Establishment of quantitative analysis model for detecting the soluble solids content in strawberry by merging near infrared spectroscopy and color parameters

CAI Deling^{1*}, TANG Chunhua², LIANG Yuying¹, ZENG Chuan¹, PENG Bining¹

1(Technical Center of Gongbei Customs District P. R., Zhuhai 519000, China) 2(Zhuhai City Polytechnic, Zhuhai 519090, China)

ABSTRACT Soluble solids content (SSC) is a key index to evaluate the quality of strawberries. In order to achieve the non-destructive evaluation of SSC, the near infrared spectroscopy was used to build the linear partial least squares (PLS) and nonlinear least squares support vector machine (LS-SVM) models. 27 effective variables were selected from the original 4 254 variables by combining both Monte-Carlo uninformative variable elimination and successive projections algorithm (MC-UVE-SPA). The quantitative analysis models were then established by using the selected effective variables. At the same time, color feature parameters were obtained based on components of RGB images of samples considering the influence of surface color on strawberries, and the multi-parameter PLS and LS-SVM models were constructed by further integrating spectra and color features. Based on the same correction set and prediction set, the prediction performance of all models for SSC was compared. The results showed that MC-UVE-SPA was an effective spectral variable selection algorithm, and the multi-parameter fusion nonlinear LS-SVM model was the optimal model for the quantitative prediction of SSC in strawberries. For samples in the prediction set, the correlation coefficient (R_p) and root mean square error of prediction (RMSEP) of the model were 0.988 5 and 0.153 2, respectively. This study lays a foundation for the development of portable instruments and online detection equipment for the detection of soluble solids content in strawberries based on near infrared spectroscopy.

Key words strawberry; near infrared spectroscopy; soluble solids content; parameter compensation; modeling analysis

第一作者:硕士,工程师(本文通讯作者, E-mail:caidelingkitty@163.com)

基金项目:中华人民共和国拱北海关科研项目(ZH2017-29)

收稿日期:2019-11-03, 改回日期:2019-12-18

水果是一类重要的农产品,是人类饮食结构的基本组成部分。2017年,我国水果总产量2.52亿t,位居世界第一位^[1-2]。果品产业已成为我国种植业中继粮食、蔬菜之后的第三大产业。水果采后分级处理对提升水果的质量和商品竞争力非常关键。多年来,水果质量快速无损检测技术一直备受关注^[3-4]。水果质量包括外部质量(如大小、颜色、形状和表皮缺陷)和内部质量(如糖度、酸度、硬度)。随着人们生活质量的提升,消费者越来越关注水果的内部质量^[5]。水果内部质量的无损检测为消费者提供口感更好的水果,从而提升行业竞争力和企业盈利能力^[6]。

水果可溶性固形物含量(soluble solids content, SSC),俗称糖度,是衡量水果内部品质最重要的指标之一^[7],该参数也决定着果实成熟度和收获时间^[8-9]。近20年来,大量的无损检测技术如近红外光谱^[6]、电子鼻^[10]、介电技术^[11]、高光谱成像技术^[12]等已经用于鲜果内部可溶性固形物含量的评估。在这些技术中,近红外光谱技术具有快速、简便、无损等特点,是目前应用最广泛的鲜果内部SSC检测方法。然而,目前大部分研究在水果内部SSC评估中没有考虑水果表皮颜色的影响,但近年来的一些研究发现,在水果内部SSC预测时,水果表皮的颜色会对预测精度产生影响^[13-14]。本研究将以小尺寸水果的典型代表草莓为研究对象,采用近红外光谱技术,融合光谱信息和颜色参数,构建多参数线性偏最小二乘模型和非线性最小二乘支持向量机模型对草莓内部可溶性固形物含量进行比较评估,同时,采用蒙特卡罗无信息变量消除结合连续投影算法对模型进行了优化,以降低模型的复杂程度和提升模型的稳健性。

1 材料与方法

1.1 实验样本及样本集划分

新鲜草莓于2019年3月采自广东珠海当地草莓园。实验中,剔除一些带有表皮缺陷的草莓,最终选择164个草莓作为研究样本,这些样本表皮颜色从红绿到全红均有分布。采集样本数据之前,需将其置于室温(20 ± 1)℃中超过24h以消除温度对预测模型精度的影响^[15]。数据采集后,所有样本被划分为校正集和预测集。具体方法为:164个样本按照可溶性固形物含量值进行从小到大排序;每4个样本中选取

第2个样本作为预测集样本,这样预测集中包含41个样本用于校正模型的评估,剩余123个草莓作为校正集样本用于校正模型的构建。在模型开发的过程中,所有模型校正集样本和预测集样本保持不变。

1.2 近红外光谱获取

使用Antaris™ II傅立叶变换近红外光谱仪(Thermo Fisher Scientific Inc., Madison, WI, USA)采集草莓样本近红外漫反射光谱。每个样本采集并获取1条光谱曲线,波数范围为 $12\,000 \sim 3\,800\text{ cm}^{-1}$,相邻波数间隔为 1.928 cm^{-1} ,因此,每条光谱曲线包含4 254个变量点。采集完光谱之后,采用Unscrambler V 9.7 software(CAMO PRECESS AS, Oslo, Norway)软件将原始反射光谱转换为吸收光谱用于随后的建模分析。

1.3 SSC和颜色参考值测量

考虑到SSC测量需要对样本进行破坏检测,因此,样本在测量之前,首先采用彩色相机(型号:AD-080GE, JAI, Japan)采集每个样本的RGB图像以用于计算每个草莓的颜色参考值(图像采集物距450mm,相机曝光时间1ms)。在图像采集完之后,立即去除样本果梗,果肉榨汁且摇晃均匀,使用数字显示手持型折射计(Model: PR-101 α , Atago Co, Ltd, Tokyo, Japan)测量果汁,3次测量并进行读数,3次读数的均值即为该样本最终SSC参考值。

每个草莓样本颜色参考值的具体计算方法步骤:

- (1)提取样本RGB图像各分量图像;
- (2)基于R分量图像和图像直方图通过阈值分割($T1 = 160$)获取仅仅带有果梗的二值化图像B1;
- (3)基于B分量图像和图像直方图通过阈值分割($T2 = 165$)获取整个草莓样本的二值化图像B2;
- (4)图像B2-B1获得二值化图像B3,进一步对B3执行形态学去噪处理,获得最终的样本果肉二值化图像B4;
- (5)采用B4对原始3个分量图像做掩模运算获得最终仅包含样本果肉信息的去除背景的样本图像,并作为目标图像;
- (6)分别计算3幅目标图像的强度均值(注:在此,图像的强度为图像灰度等级,级别为256级,灰度等级越高图像越亮),该均值作为衡量样本颜色信息的参考值(图1)。阈值T1和T2是基于对所有校正集样本分析所获得。

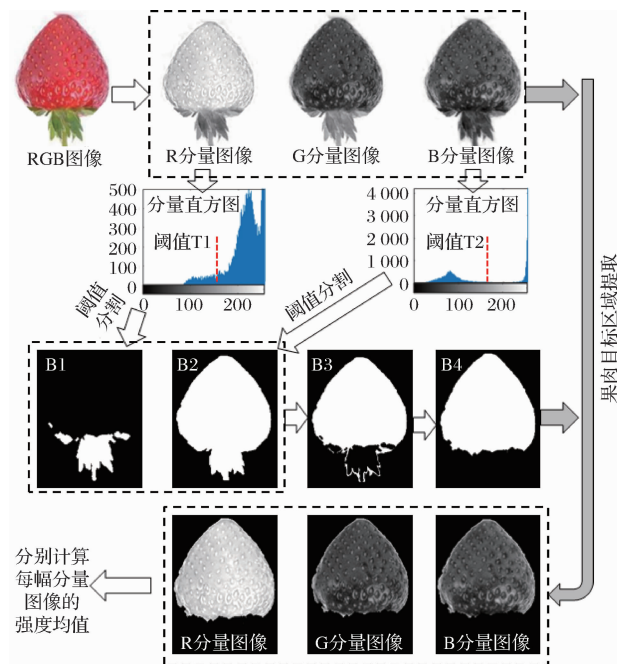


图1 样本颜色参考值获取流程

Fig. 1 The flow chart of acquisition of color reference value of samples

1.4 变量选择方法

在原始全谱中,一些无信息的变量不包含与草莓可溶性固形物分析相关的信息,因此,消除这些变量可以提高模型的预测能力。无信息性变量消除(uninformative variable elimination, UVE)是一种非常有效的变量选择方法,其目标是去除建模信息比噪声少的变量^[16]。在 UVE 算法中,通过留一交互验证计算偏最小二乘(partial least squares, PLS)回归系数矩阵 $b = [b_1, \dots, b_p]$, 然后根据其稳定性对每个变量的可靠性进行测量。变量 j 的稳定性可按公式(1)计算:

$$S_j = \text{mean}(\beta_j) / \text{std}(\beta_j) \quad (1)$$

式中: $\text{mean}(\beta_j)$ 和 $\text{std}(\beta_j)$ 分别代表变量 j 回归系数的均值和标准差。绝对稳定性越大,对应的变量越重要。在截止阈值之间的变量认为是无信息性的,并被剔除。为了确定无信息的变量,随机变量需附加到数据集以确定截止阈值。在蒙特卡罗无信息性变量消除(Monto-Carlo UVE, MC-UVE)算法中,将蒙特卡罗策略引入 UVE-PLS 中,代替了传统的留一交互策略。此外,变量直接由稳定性来确定,这比 UVE 方法中在原始数据矩阵中加入随机噪声变量来估计截止阈值更方便。MC-UVE 算法参考文献[17]。

除了无信息变量外,原始光谱中还可能存在共线或冗余变量。连续投影算法(succesive projections al-

gorithm, SPA)算法的目标是选取冗余最小的变量来解决共线性问题^[18]。SPA 是一种正向变量选择算法,它从一个变量开始,然后在每次迭代中加入一个新的变量,直到达到指定数量的 N 个变量。通过 SPA 选择有效变量,步骤如下:

- (1) 所选特征波数记为 N , 参考向量记为 $k(n-1)$, 第 1 次计算时 $n=1$;
- (2) 初始选择波数为 $k(0)$, 从校正集光谱 X_{cal} (I 行) 样品数; J 列(波数变量数; $N < J$) 中选择第 j 列波数变量列向量 x_j ($j=1 \dots J$) 做参考向量;
- (3) 剩余列向量记为 S , $S = \{j, 1 \leq j \leq J, \text{且 } j \in / \{k(0) \dots k(n-1)\}\}$;
- (4) 计算 x_j 在 S 中各列向量正交子空间上的投影算子 $P_{x_j}(j \in S)$: $P_{x_j} = x_j - (x_j^T x_{k(n-1)}) x_{k(n-1)} (x_{k(n-1)}^T x_{k(n-1)})^{-1}$;
- (5) 若 $n < N$, 记 $n = n + 1$, 标记 S 中 x_j 的最大投影算子所在列向量为新参考向量, 重复步骤(2)和(3);

(6) 结果为所选变量 $X_s = \{k(n-1); n=1 \dots N\}$, 是光谱的最小共线性子矩阵;

(7) 计算初始波数 $k(0)$ 时, 校正集子光谱 $X_{\text{cal}}(X_s)$ 与校正集样品 SSC 值 Y_{cal} 建立 MLR 模型, 计算该模型验证集子光谱 $X_{\text{val}}(X_s)$ 预测标准偏差(root mean square error of prediction, RMSEP);

(8) 重复上述(2)~(5)步骤, 计算不同 $k(0)$ 时模型 RMSEP, 最小 RMSEP 值对应 $k(0)$ 的 X_i 最佳;

(9) $N = N + 1$, 重复上述步骤, 确定 $N = N + 1$ 时最佳 X_i ;

(10) 不同 N 值对应最佳 X_i 建立模型的 RMSEP 的最小值对应 N 值为最佳变量数, 对应 X_i 为所选特征波数变量。

1.5 模型构建和评估

本研究构建两类模型即线性偏最小二乘模型和非线性最小二乘支持向量机模型(least squares-support vector machine, LS-SVM)用于草莓可溶性固形物含量预测。

偏最小二乘(PLS)是目前水果内部可溶性固形物含量中广泛使用的线性建模方法。建模时, PLS 同时考虑了 SSC 值和输入变量数据, 找出两者之间的基本关系。本研究在模型的开发过程中, 采用全交叉验证法, 通过交叉验证的均方根误差(root mean square error of calibration, RMSEC)来确定潜在变量 LV 的最优数目, 以避免模型过拟合问题。

支持向量机是一种典型的统计学习方法,它通过结构风险最小化原理来提高泛化能力,较好地解决了小样本、非线性、高维数、局部极小等实际问题。最小二乘支持向量机(LS-SVM)是其改进版本,是一种先进的学习算法,LS-SVM 能够处理线性和非线性多变量问题,并以一种相对快速的方式解决这些关系^[19]。LS-SVM 算法参考文献[20]。

所有模型的预测性能通过相关系数(R)、RM-SEC、RMSEP 进行评估。评估参数计算公式:

$$R_c, R_p = \sqrt{1 - [\sum_{i=1}^n (y_i - \hat{y}_i)^2 / \sum_{i=1}^n (y_i - y_m)^2]} \quad (2)$$

$$\sum_{i=1}^n \text{RMSEC}, \text{RMSEP} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

式中: y_i 和 \hat{y}_i 代表在校正集或者预测集中第 i 个样本的实测 SSC 值和预测 SSC 值; y_m 代表在校正集或预测集中样本实测 SSC 的均值; n 代表校正集或预测集中的样本数。一个好的模型通常具有低的 RMSEC 和 RMSEP 值,高的 R_c 和 R_p 值^[3]。

2 结果与讨论

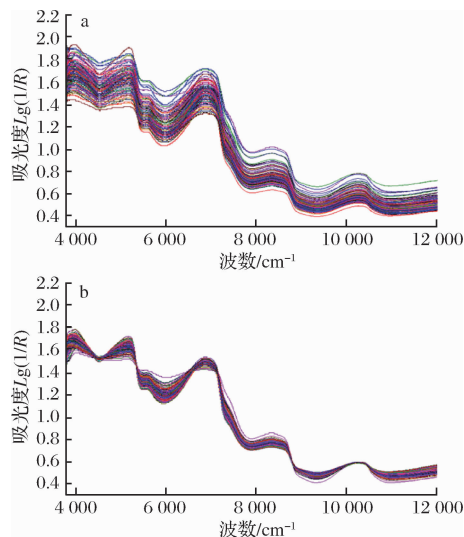
2.1 光谱分析

图2表示草莓样本原始光谱和通过多元散射校正预处理后的光谱。通过光谱预处理,原始光谱所存在的散射得到了有效的校正。从光谱图中可以看到,所有样本光谱曲线类似,不存在明显的异常样本,并且光谱波数较大时,光谱吸收强度更大,主要原因在于波数较大区域存在着明显的 H_2O 吸收(如位于波数 $6\,944\text{ cm}^{-1}$ 和 $5\,155\text{ cm}^{-1}$ 的吸收峰)^[21]。另外,光谱曲线中也存在一些小的吸收峰如 $8\,403\text{ cm}^{-1}$,这些吸收峰与C—H二级倍频有关系。所有这些吸收特性均有助于草莓内部 SSC 的预测。

2.2 参考值统计结果

草莓样本不同数据集可溶性固形物含量和颜色参考值统计结果见表1。

从表1可看出,就 SSC 含量而言,校正集中样本 SSC 值范围为 $6.18 \sim 13.57\text{ }^\circ\text{Brix}$,预测集为 $6.50 \sim 13.10\text{ }^\circ\text{Brix}$,前者范围包含后者。对于样本颜色参考值,R、G 和 B 分量图像的强度值在校正集和预测集的范围分别为 $173.64 \sim 217.12$ 和 $174.25 \sim 215.96$ 、 $72.67 \sim 90.21$ 和 $74.59 \sim 88.11$ 、 $77.32 \sim 95.66$ 和 $78.56 \sim 93.21$ 。对每个分量,校正集样本强度值范围涵盖了预测集样本强度值范围。这些特征有助于构建一个相对稳健的预测模型。



a-草莓样本原始光谱;b-草莓样本预处理后光谱图

图2 草莓样本原始光谱和预处理光谱

Fig. 2 Raw and preprocessing spectra of strawberry samples

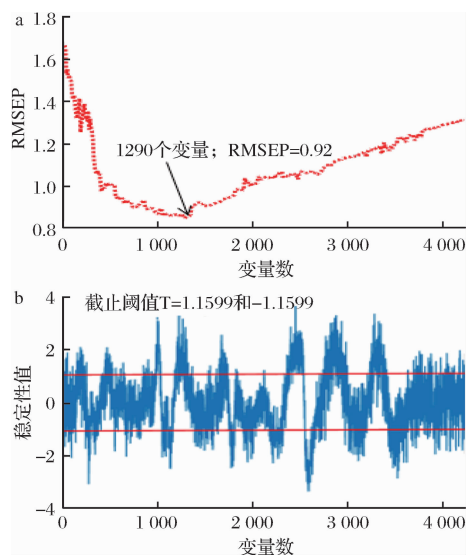
表1 草莓样本不同数据集可溶性固形物含量和颜色参考值统计

Table 1 Statistical values of SSC and color parameters of strawberry samples in different data sets

参数类型	数据集	最大值	最小值	均值	标准差
SSC/ $^\circ\text{Brix}$	总样本集	13.57	6.18	10.09	1.02
	校正集	13.57	6.18	10.39	1.11
	预测集	13.10	6.50	10.25	1.07
R 分量	总样本集	217.12	173.64	197.26	10.20
	校正集	217.12	173.64	195.77	10.12
	预测集	215.96	174.25	193.74	9.96
G 分量	总样本集	90.21	72.67	81.85	10.36
	校正集	90.21	72.67	80.63	9.88
	预测集	88.11	74.59	79.52	10.06
B 分量	总样本集	95.66	77.32	87.16	10.11
	校正集	95.66	77.32	85.92	11.31
	预测集	93.21	78.56	84.27	10.41

2.3 有效变量选择结果

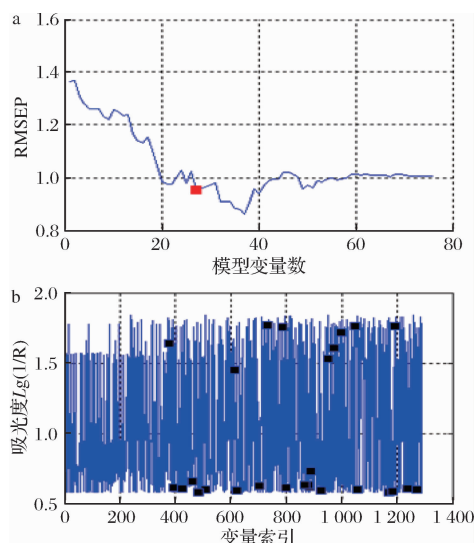
使用 MC-UVE 和 SPA 联立算法即 MC-UVE-SPA 从全光谱 4 254 个变量中选择有效变量。在 MC-UVE 变量选择过程中,利用校正集中样本的光谱数据和 SSC 参考值作为输入,通过研究不同 PLS 校准模型预测 RMSEP 的变化确定最优变量数。以 RMSEP 最低对应的变量数作为最优。变量选择结果如图3所示,由图3-a可以看出,选取1 290个变量时所建模型最优。这1 290个有效变量可通过评估图3-b所示的变量稳定性来确定。图3-b中的2条水平线代表上下截止阈值。截止线以内的变量被视为无信息变量。



a-SSC 分析模型 RMSEP 曲线图;b-变量稳定性评估图
图3 蒙特卡罗无信息变量消除法对变量进行选择的结果

Fig.3 Effective variable selection by MC-UVE

使用 MC-UVE 选择变量后,剔除了原始光谱数据中的无信息变量。然后在 MC-UVE 之后进行 SPA 变量选择,通过消除冗余变量来提高预测性能。因此,使用 1 290 个变量作为 SPA 的输入。在 SPA 变量选择过程中,计算所有变量子集所建模型的 RMSEP 值,图 4-a 中红色实心方框代表最优的变量数,即 27 个变量被确认为有效变量。这 27 个变量对应 1 290 个变量的索引如图 4-b 所示。从图 4-b 中可以看出,变量数被极大地压缩,且变量之间的共线性明显消除。



a-RMSEP 值变化曲线;b-变量索引图

图4 连续投影算法有效变量选择结果

Fig.4 Effective variable selection by SPA

2.4 模型预测结果分析

3 组数据包括全光谱变量(4 254 个)、有效光谱变量(27 个)和融合变量(30 个,即 27 个有效光谱变量+3 个颜色参数),分别作为输入构建 PLS 和 LS-SVM 模型用于草莓样本可溶性固形物含量预测。PLS 模型预测结果显示见表 2。从表 2 可看出,由于全变量参与建模,太多的变量引起了 PLS 模型过拟合,从而导致模型的预测性能大大降低(此现象也发生在表 3 中的全变量 LS-SVM 模型中),通过变量选择,PLS 模型的预测性能得到了明显提升, R_p 由原来的 0.752 3 提高到 0.947 0, RMSEP 由原来的 0.862 1 降为 0.381 1。然而,从表 2 中也可看出,颜色参数的加入,并没有使 MC-UVE-SPA-PLS 模型的预测性能得到提高,这表明颜色参数与草莓可溶性固形物之间可能不存在线性关系。

表2 基于不同输入构建的 PLS 模型对草莓 SSC 含量预测结果

Table 2 Prediction results for SSC in strawberry by PLS models developed based on different inputs

模型类型	变量数	LVs	校正集		预测集	
			R_c	RMSEC	R_p	RMSEP
Full-spectrum-PLS	4 254	15	0.954 2	0.344 2	0.752 3	0.862 1
MC-UVE-SPA-PLS	27	9	0.9411	0.387 1	0.947 0	0.381 1
MC-UVE-SPA-PLS + 颜色参数	30	10	0.941 2	0.387 0	0.947 3	0.380 9

在 LS-SVM 建模过程中,需要确定模型参数 γ 和 σ^2 ,在此采用两步网格搜索技术进行参数寻优。作为示例,图 5 显示了 MC-UVE-SPA-PLS + 颜色参数模型参数寻优图。网格点“●”和“×”分别代表第一步和第二步网格搜索的范围和步长。曲线表示轮廓误差。初始值的选择对结果没有影响,在这项研究中, γ 和 σ^2 的初始值设置为 0.01。对于 3 类不同的 LS-SVM 模型, γ 和 σ^2 的值分别为 $(32, 2.5 \times 10^4)$ 、 $(31, 7.0 \times 10^3)$ 和 $(2.8 \times 10^3, 6.9 \times 10^2)$ 。

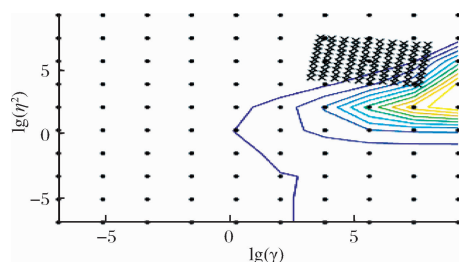


图5 MC-UVE-SPA-PLS + 颜色参数模型参数优化

Fig.5 Optimization of γ and σ^2 for MC-UVE-SPA-LS-SVM + color parameter model

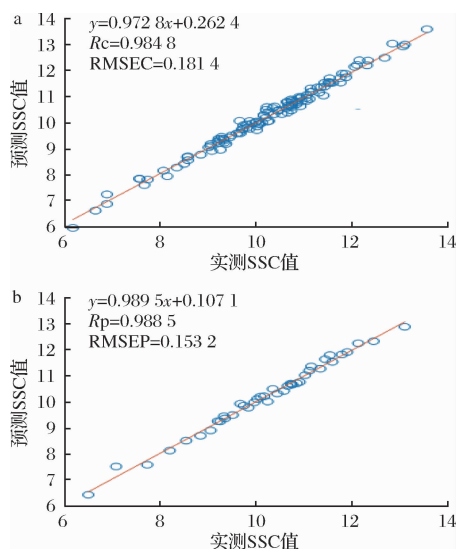
基于 3 类 LS-SVM 模型对草莓 SSC 进行预测,结果如表 3 所示。

表 3 基于不同输入构建的 LS-SVM 模型对草莓 SSC 含量预测结果

Table 3 Prediction results for SSC in strawberry by LS-SVM models developed based on different inputs

模型类型	变量数	gamma, sig ²	校正集		预测集	
			R _c	RMSEC	R _p	RMSEP
Full-spectrum-LS-SVM	4 254	32, 2.5 × 10 ⁴	0.936 5	0.401 2	0.731 9	0.896 3
MC-UVE-SPA-LS-SVM	27	31, 7.0 × 10 ³	0.931 1	0.423 6	0.932 9	0.421 1
MC-UVE-SPA-LS-SVM + 颜色参数	30	2.8 × 10 ³ , 6.9 × 10 ²	0.984 8	0.181 4	0.988 5	0.153 2

从表 3 中可以看出,类似 PLS 模型,通过变量选择后,LS-SVM 模型的性能有了明显提升,表明变量选择的重要性。与 MC-UVE-SPA-PLS 模型相比,MC-UVE-SPA-LS-SVM 模型有着类似或者略低的预测精度。但当颜色参数和光谱同时作为输入时,所构建的 LS-SVM 模型(MC-UVE-SPA-LS-SVM + 颜色参数)预测性能明显好于 MC-UVE-SPA-LS-SVM 和 MC-UVE-SPA-PLS 模型,这表明草莓颜色和 SSC 之间存在着某种非线性关系,这种非线性关系有助于草莓 SSC 的准确评估。因此,通过比较发现,融合光谱和颜色参数所构建的 LS-SVM 模型最优,基于该模型对所有样本进行预测,图 6-a 显示了校正集样本的预测散点图,图 6-b 预测集样本的预测散点图,从图中可以看出,样品分布在回归曲线附近,且接近回归曲线,说明 MC-UVE-SPA-LS-SVM + 颜色参数模型在草莓内部可溶性固形物评估中具良好性能。



a-校正集样本的预测散点图;b-预测集样本的预测散点图

图 6 最优的 LS-SVM 模型预测结果散点图

Fig.6 Scatter plots of predicted versus measured SSC obtained by the best LS-SVM model

本文研究表明,多参数融合可能更有助于草莓内

部可溶性固形物含量 SSC 的预测,该研究结论与 WANG 等^[13]和 GUO 等^[14]在苹果和西红柿内部 SSC 检测评估中所获得的结果类似。

3 结论

本研究采用近红外光谱技术结合颜色参数补偿建模方法和变量选择算法实现了草莓内部可溶性固形物含量的准确预测。比较了不同输入下所构建的 PLS 模型和 LS-SVM 模型的预测性能,一方面发现基于有效变量(仅为全光谱变量的 0.63%)所构建的预测模型(MC-UVE-SPA-PLS 和 MC-UVE-SPA-LS-SVM)对草莓可溶性固形物评估性能明显优于全光谱变量模型(Full-spectrum-PLS 和 Full-spectrum-LS-SVM),表明本研究所建议的 MC-UVE-SPA 联合变量选择算法的有效性;另一方面也发现,融合光谱变量和颜色参数所构建的非线性 LS-SVM 模型预测性能优于线性 MC-UVE-SPA-PLS 模型,表明草莓颜色与其 SSC 之间有着某种非线性的关系,并且这种非线性关系有助于提升草莓内部 SSC 的预测精度。针对建模集和预测集草莓样本,基于融合光谱变量和颜色参数的最优模型 MC-UVE-SPA-LS-SVM 对 SSC 预测精度 R_c 和 RMSEC 分别为 0.984 8 和 0.181 4, R_p 和 RMSEP 分别为 0.988 5 和 0.153 2。

参 考 文 献

- [1] 国家统计局. 中国统计年鉴[M]. 北京: 中国统计出版社, 2018.
- [2] FAOSTAT Statistical Database: Food and Agriculture Organization of the United Nations. <http://www.fao.org/faostat>.
- [3] NICOLAÏ B M, BEULLENS K, BOBELYN E, et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review[J]. Postharvest Biology and Technology, 2007, 46: 99–118.
- [4] ZHANG B H, DAI D J, HUANG J C, et al. Influence of physical and biological variability and solution methods in fruit and vegetable quality nondestructive inspection by using imaging and near-infrared spectroscopy techniques: A review[J]. Critical Reviews in Food Science and Nutrition, 2017, 58(12): 2 099–2 118.

- [5] PARK E, LUO Y G, MARINE S C, et al. Consumer preference and physicochemical evaluation of organically grown melons [J]. *Postharvest Biology and Technology*, 2018, 141: 77–85.
- [6] XIE L J, WANG A C, XU H R, et al. Applications of near-infrared systems for quality evaluation of fruits: A Review [J]. *Transactions of the ASABE*, 2016, 59(2): 399–419.
- [7] LI J L, SUN D W, CHENG J H. Recent advances in nondestructive analytical techniques for determining the total soluble solids in fruits: a review [J]. *Comprehensive Reviews in Food Science and Food Safety*, 2016, 15: 897–911.
- [8] RAJKUMAR P, WANG N, EIMASRY G, et al. Studies on banana fruit quality and maturity stages using hyperspectral imaging [J]. *Journal of Food Engineering*, 2012, 108: 194–200.
- [9] KOUR R, SINGH M, GILL P P S, et al. Ripening quality of *Dusehri mango* in relation to harvest time [J]. *Journal of Food Science and Technology*, 2018, 55(7): 2 395–2 400.
- [10] HUI G H, WU Y L, YE D D, et al. Study of peach freshness predictive method based on electronic nose [J]. *Food Control*, 2012, 28(1): 25–32.
- [11] GUO W C, FANG L J, LIU D Y, et al. Determination of soluble solids content and firmness of pears during ripening by using dielectric spectroscopy [J]. *Computers and Electronics in Agriculture*, 2015, 117: 226–233.
- [12] LI J B, CHEN L P. Comparative analysis of models for robust and accurate evaluation of soluble solids content in ‘Pinggu’ peaches by hyperspectral imaging [J]. *Computers and Electronics in Agriculture*, 2017, 142: 524–535.
- [13] WANG H T, ZHANG R Y, PENG Z, et al. Measurement of SSC in processing tomatoes (*Lycopersicon esculentum* Mill.) by applying Vis-NIR hyperspectral transmittance imaging and multi-parameter compensation models [J]. *Journal of Food Process Engineering*, 2019, 42(5): e13100.
- [14] GUO Z M, HUANG W Q, PENG Y K, et al. Color compensation and comparison of shortwave near infrared and long wave near infrared spectroscopy for determination of soluble solids content of ‘Fuji’ apple [J]. *Postharvest Biology and Technology*, 2016, 115: 81–90.
- [15] COZZOLINO D, LIU L, CYNKAR W U, et al. Effect of temperature variation on the visible and near infrared spectra of wine and the consequences on the partial least square calibrations developed to measure chemical composition [J]. *Analytica Chimica Acta*, 2007, 588(2): 224–230.
- [16] WU D, HE Y, NIE P C, et al. Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice [J]. *Analytica Chimica Acta*, 2010, 659: 229–237.
- [17] CAI W S, LI Y K, SHAO X G. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra [J]. *Chemometrics and Intelligent Laboratory Systems*, 2008, 90: 188–194.
- [18] ARAÚJO M C U, SALDANHA T C B, GALVÃO R K H, et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis [J]. *Chemometrics and Intelligent Laboratory Systems*, 2001, 57: 65–73.
- [19] SUYKENS J A K, VANDEWALLE J. Least squares support vector machine classifiers [J]. *Neural Processing Letters*, 1999, 9(3): 293–300.
- [20] SUYKENS J A K, VAN GESTEL T, DE BRABANTER J, et al. Least squares support vector machines [J]. *World Scientific*, Singapore, 2002.
- [21] LI J B, HUANG W Q, CHEN L P, et al. Variable selection in visible and near-infrared spectral analysis for noninvasive determination of soluble solids content of ‘Ya’ pear [J]. *Food Anal. Methods*, 2014, 7: 1 891–1 902